

# THE THEOREM OF HALMOS AND SAVAGE UNDER FINITE ADDITIVITY

GIANLUCA CASSESE

**ABSTRACT.** Given a generalization of Lebesgue decomposition we obtain an extension to the finitely additive setting of the theorems of Halmos and Savage and of Yan.

## 1. INTRODUCTION AND NOTATION

In a paper that soon became a classic in statistics [7], Halmos and Savage illustrated the powerful implications of the Radon Nikodym theorem for the theory of sufficient statistics. One of their results, Lemma 7, deals with dominated sets of probability measures and states that each such set admits an equivalent, countable subset. This lemma rapidly obtained its own popularity, proving to be very useful in a variety of different contexts, such as the proof of Yan Theorem, another classical result in probability and in mathematical finance.

In their proof, Halmos and Savage exploit extensively countable additivity and the fact that the underlying family is a  $\sigma$ -algebra. Both properties are essential as they allow, loosely speaking, for the possibility of taking limits. For this reason their method of proof cannot be adapted to the case in which probability is just *finitely* additive, a situation of interest for the subjective theory of probability originating from the seminal work of de Finetti [4] and, more generally, for decision theory in which countable additivity is more an exception than a rule. Finite additivity is also unavoidable in many classical problems in which it is needed to take extensions of the given set function.

In this short note we extend the original result of Halmos and Savage to the case of finitely additive probability measures and obtain, as a corollary, an analogous extension of the theorem of Yan [10]. The proof is, somehow surprisingly, straightforward and does not make use but of classical decomposition results of set functions, ultimately due to Bochner and Phillips.

In the following,  $\Omega$  will be a fixed, nonempty set and  $\mathcal{A}$  an algebra of subsets of  $\Omega$ . Also given is a positive, additive, bounded set function  $\lambda \in ba(\mathcal{A})_+$ . A set  $\mathcal{M} \subset ba(\mathcal{A})$  is said to be dominated by  $\lambda$  if  $\mu \ll \lambda$  for every  $\mu \in \mathcal{M}$  (in symbols  $\mathcal{M} \ll \lambda$ ). For the theory of finitely additive measures and integrals we mainly borrow notation, definitions and terminology from Dunford and Schwartz [6], although we prefer the symbol  $|\mu|$  to denote the total variation measure generated by  $\mu$  and we

---

*Date:* January 31, 2014.

*2000 Mathematics Subject Classification.* Primary 28A33, Secondary 46E27.

*Key words and phrases.* Lebesgue decomposition, Halmos Savage Theorem, Yan Theorem.

write  $\mu_f$  to denote that element of  $ba(\mathcal{A})$  defined implicitly by letting

$$(1) \quad \mu_f(A) = \int \mathbf{1}_A f d\mu \quad A \in \mathcal{A}$$

whenever  $f \in L^1(\mu)$ . We often write  $\mu(f)$  rather than  $\int f d\mu$ .

The lattice symbol  $\perp$  is used to define the orthogonal complement

$$\mathcal{M}^\perp = \{\nu \in ba(\mathcal{A}) : \nu \perp \mu \text{ for every } \mu \in \mathcal{M}\}$$

of  $\mathcal{M}$  which is known to be a normal sublattice of  $ba(\mathcal{A})$ , see e.g. [1, 1.5.6 and 1.5.8].

## 2. A DECOMPOSITION

We associate with  $\mathcal{M} \subset ba(\mathcal{A})$  the collections

$$(2) \quad \mathbf{A}(\mathcal{M}) = \left\{ \sum_n \alpha_n \frac{|\mu_n|}{1 \vee \|\mu_n\|} : \mu_n \in \mathcal{M}, \alpha_n \geq 0 \text{ for } n = 1, 2, \dots, \sum_n \alpha_n = 1 \right\}$$

$$(3) \quad \mathbf{L}(\mathcal{M}) = \{\nu \in ba(\mathcal{A}) : \nu \ll m \text{ for some } m \in \mathbf{A}(\mathcal{M})\}$$

To obtain a simple generalization of Lebesgue decomposition, we start remarking that  $\mathbf{L}(\mathcal{M})$  is a normal sublattice of  $ba(\mathcal{A})$  and so that, by Riesz decomposition Theorem [1, 1.5.10],  $ba(\mathcal{A}) = \mathbf{L}(\mathcal{M}) + \mathbf{L}(\mathcal{M})^\perp$ . To see this, take an increasing net  $\langle \nu_\alpha \rangle_{\alpha \in \mathfrak{A}}$  in  $\mathbf{L}(\mathcal{M})$  with  $\nu = \lim_\alpha \nu_\alpha \in ba(\mathcal{A})$ , extract a sequence  $\langle \nu_{\alpha_n} \rangle_{n \in \mathbb{N}}$  such that  $\|\nu - \nu_{\alpha_n}\| = (\nu - \nu_{\alpha_n})(\Omega) < 2^{-n-1}$ , choose  $m_n \in \mathbf{A}(\mathcal{M})$  such that  $m_n \gg \nu_{\alpha_n}$  and define  $m = \sum_n 2^{-n} m_n \in \mathbf{A}(\mathcal{M})$ . Since  $m \gg m_n \gg \nu_{\alpha_n}$  for each  $n \in \mathbb{N}$ , there is  $\delta_n > 0$  such that  $m(A) < \delta_n$  implies  $|\nu_{\alpha_n}|(A) < 2^{-n-1}$  and, therefore,  $|\nu|(A) \leq |\nu_{\alpha_n}|(A) + 2^{-n-1} \leq 2^{-n}$ . This proves that if  $\{\nu_\alpha : \alpha \in \mathfrak{A}\}$  is a nonempty family in  $\mathbf{L}(\mathcal{M})$  and if  $\bigvee_{\alpha \in \mathfrak{A}} \nu_\alpha$  exists in  $ba(\mathcal{A})$ , then necessarily  $\bigvee_{\alpha \in \mathfrak{A}} \nu_\alpha \in \mathbf{L}(\mathcal{M})$ . Moreover,  $|\nu_1| \leq |\nu|$  and  $\nu \in \mathbf{L}(\mathcal{M})$  imply  $\nu_1 \in \mathbf{L}(\mathcal{M})$ . Noting that  $\mathbf{L}(\mathcal{M})^\perp = \mathbf{A}(\mathcal{M})^\perp$  we obtain the following:

**Lemma 1.** *For each  $\lambda \in ba(\mathcal{A})$  and  $\mathcal{M} \subset ba(\mathcal{A})$  there is a unique way of writing*

$$(4) \quad \lambda = \lambda_{\mathcal{M}}^c + \lambda_{\mathcal{M}}^\perp$$

*with  $\lambda_{\mathcal{M}}^c \in \mathbf{L}(\mathcal{M})$  and  $\lambda_{\mathcal{M}}^\perp \perp \mathbf{A}(\mathcal{M})$ . If  $\lambda$  is positive or countably additive then so are  $\lambda_{\mathcal{M}}^\perp$  and  $\lambda_{\mathcal{M}}^c$ .*

## 3. THE HALMOS-SAVAGE THEOREM AND ITS IMPLICATIONS

We now prove the main result of the paper. Let us mention that dominated sets of measures arise whenever dealing with a model, a statistical model e.g., in which it is posited the existence of a reference probability measure.

**Theorem 1** (Halmos and Savage).  *$\mathcal{M} \subset ba(\mathcal{A})$  is dominated if and only if  $\mathcal{M} \ll m$  for some  $m \in \mathbf{A}(\mathcal{M})$ .*

*Proof.*  $\lambda$  dominates  $\mathcal{M}$  if and only if  $\lambda_{\mathcal{M}}^c$  does. In fact, choose  $\mu \in \mathcal{M}$  and  $\varepsilon > 0$  and let  $\delta$  be such that  $\lambda(A) < \delta$  implies  $|\mu|(A) < \varepsilon$ . Pick  $B \in \mathcal{A}$  such that  $|\mu|(B^c) + \lambda_{\mathcal{M}}^{\perp}(B) < (\delta/2) \wedge \varepsilon$ . Then  $\lambda_{\mathcal{M}}^c(A) < \delta/2$  implies  $\lambda(A \cap B) < \delta$  and thus  $|\mu|(A) \leq |\mu|(A \cap B) + \varepsilon \leq 2\varepsilon$ . Lemma 1 proves the claim.  $\square$

To rephrase the above Theorem in the language of Halmos and Savage, observe that if  $\mathcal{M}_0 = \{\mu_1, \mu_2, \dots\}$  is the subfamily of  $\mathcal{M}$  generating  $m = \sum_n 2^{-n} |\mu_n| / (1 \vee \|\mu_n\|)$  and  $\langle A_k \rangle_{k \in \mathbb{N}}$  is a sequence in  $\mathcal{A}$ , then  $\lim_k |\mu_n|(A_k) = 0$  for  $n = 1, 2, \dots$  if and only if  $\lim_k |\mu|(A_k) = 0$  for all  $\mu \in \mathcal{M}$  and  $\mathcal{M}_0$  may then be said to be *equivalent* to  $\mathcal{M}$ .

A typical application is the following:

**Corollary 1.** *Let  $\mathcal{H} \subset \mathcal{A}$  and  $\mathcal{A}_{\mathcal{H}} = \{A \in \mathcal{A} : \inf_{\{\alpha \subset \mathcal{H}, \alpha \text{ finite}\}} \lambda(A \setminus \bigcup_{\alpha} H) = 0\}$ . There exists  $H_1, H_2, \dots \in \mathcal{H}$  such that*

$$(5) \quad \lim_k \sup_{A \in \mathcal{A}_{\mathcal{H}}} \lambda \left( A \setminus \bigcup_{n \leq k} H_n \right) = 0$$

*Proof.* Write  $\mathcal{M} = \{\lambda_H : H \in \mathcal{H}\}$  and choose  $m = \sum_n \alpha_n \lambda_{H_n} \in \mathbf{A}(\mathcal{M})$  to be such that  $m \gg \mathcal{M}$ . By construction, for each  $H \in \mathcal{H}$ , we conclude  $\lim_k \lambda(H \setminus \bigcup_{n=1}^k H_n) = \lim_k \lambda_H(\bigcap_{n=1}^k H_n^c) = 0$ . Consider a disjoint union  $B = \bigcup_{j=1}^I A_j \cap K_j$  with  $A_j \in \mathcal{A}$  and  $K_j \in \mathcal{H}$  for  $j = 1, \dots, I$  and denote by  $\mathcal{H}_1$  the corresponding class. But then, since  $\lambda_{K_j} \in \mathcal{M}$  for  $j = 1, \dots, I$ ,

$$\begin{aligned} \lambda \left( B \cap \bigcap_{n \leq k} H_n^c \right) &= \sum_{j=1}^I \lambda \left( A_j \cap K_j \cap \bigcap_{n \leq k} H_n^c \right) \\ &= \lim_r \sum_{j=1}^I \lambda \left( A_j \cap K_j \cap \bigcap_{n \leq k} H_n^c \cap \bigcup_{n \leq r} H_n \right) \\ &\leq \lim_r \lambda \left( \bigcap_{n \leq k} H_n^c \cap \bigcup_{n \leq r} H_n \right) \end{aligned}$$

so that  $\lim_k \sup_{B \in \mathcal{H}_1} \lambda(B \cap \bigcap_{n \leq k} H_n^c) = 0$ . Let now  $A \in \mathcal{A}_{\mathcal{H}}$ . We have

$$\begin{aligned} \lim_k \sup_{A \in \mathcal{A}_{\mathcal{H}}} \lambda \left( A \cap \bigcap_{n \leq k} H_n^c \right) &= \lim_k \sup_{A \in \mathcal{A}_{\mathcal{H}}} \sup_{K_1, \dots, K_I \in \mathcal{H}} \lambda \left( A \cap \bigcup_{j=1}^I K_j \cap \bigcap_{n \leq k} H_n^c \right) \\ &\leq \lim_k \sup_{B \in \mathcal{H}_1} \lambda \left( B \cap \bigcap_{n \leq k} H_n^c \right) \\ &= 0 \end{aligned}$$

which proves (5).  $\square$

For the next result, define the  $\lambda$ -completion of  $\mathcal{A}$  as follows

$$(6) \quad \mathcal{A}(\lambda) = \left\{ B \subset \Omega : \inf_{\{A, A' \in \mathcal{A} : A \subset B \subset A'\}} \lambda(A' \setminus A) = 0 \right\}$$

It is clear that  $\lambda$  admits exactly one extension to  $\mathcal{A}(\lambda)$  defined by letting

$$(7) \quad \bar{\lambda}(B) = \sup_{\{A \in \mathcal{A} : A \subset B\}} \lambda(A) = \inf_{\{A' \in \mathcal{A} : B \subset A'\}} \lambda(A') \quad B \in \mathcal{A}(\lambda)$$

Finite additivity often emerges upon taking extensions of a countably additive set function. The following Corollary examines one such situation and establishes countable additivity holds at least locally along some sequence.

**Corollary 2.** *Let  $\mathcal{B}(\lambda) = \mathcal{A}(\lambda) \setminus \mathcal{A}$  be non empty. There exists a disjoint sequence  $\langle A_n \rangle_{n \in \mathbb{N}}$  in  $\mathcal{A}$  such that  $\bigcup_n A_n \in \mathcal{A}(\lambda)$  and*

$$(8) \quad \bar{\lambda}(B) = \sum_n \bar{\lambda}(B \cap A_n) \quad B \in \mathcal{A}(\lambda)$$

*Proof.* Choose

$$\mathcal{H} = \{H \in \mathcal{A} : H \subset B \text{ for some } B \in \mathcal{B}(\lambda)\}$$

in Corollary 1. Then  $\mathcal{B}(\lambda) \subset \mathcal{A}_{\mathcal{H}}$ . Extract the sequence  $\langle A_n \rangle_{n \in \mathbb{N}}$  from the sequence  $\langle H_n \rangle_{n \in \mathbb{N}}$  of Corollary 1 by letting  $A_n = H_n \setminus \bigcup_{j < n} H_j$  and observe that  $A_n \in \mathcal{H}$ . By (5) we obtain that  $\bar{\lambda}(B) = \sum_n \bar{\lambda}(B \cap A_n)$  for each  $B \in \mathcal{B}(\lambda)$ . Observe that  $\mathcal{B}(\lambda)$  is closed with respect to complementation and thus

$$(9) \quad \inf_{\{A' \in \mathcal{A} : \bigcup_n A_n \subset A'\}} \lambda(A') \leq \lambda(\Omega) = \bar{\lambda}(B) + \bar{\lambda}(B^c) = \sum_n \lambda(A_n) \leq \sup_{\{A \in \mathcal{A} : A \subset \bigcup_n A_n\}} \lambda(A)$$

which proves that  $\bigcup_n A_n \in \mathcal{A}(\lambda)$ . But then  $\bar{\lambda}(B) \geq \bar{\lambda}(B \cap \bigcup_n A_n) \geq \sum_n \bar{\lambda}(B \cap A_n)$  for each  $B \in \mathcal{B}(\lambda)$ . Applying this conclusion to  $B \in \mathcal{B}(\lambda)$  and its complement and exploiting (9) one concludes that (8) necessarily holds.  $\square$

Another possible development of Theorem 1 is the following finitely additive version of a theorem of Yan [10, Theorem 2, p. 220] which is well known in stochastic analysis and mathematical finance:

**Corollary 3 (Yan).** *Let  $\mathcal{K} \subset L^1(\lambda)$  be convex with  $0 \in \mathcal{K}$ , write  $\mathcal{C} = \mathcal{K} - \mathcal{S}(\mathcal{A})_+$  and denote by  $\bar{\mathcal{C}}$  the closure of  $\mathcal{C}$  in  $L^1(\lambda)$ . The following are equivalent:*

- (i) *for each  $f \in L^1(\lambda)_+$  with  $\lambda(f) > 0$  there exists  $\eta > 0$  such that  $\eta f \notin \bar{\mathcal{C}}$ ;*
- (ii) *for each  $A \in \mathcal{A}$  with  $\lambda(A) > 0$  there exists  $d > 0$  such that  $d\mathbf{1}_A \notin \bar{\mathcal{C}}$ ;*
- (iii) *there exists a finitely additive probability  $P$  on  $\mathcal{A}$  such that*
  - (a)  $\mathcal{K} \subset L^1(P)$  and  $\sup_{k \in \mathcal{K}} P(k) < \infty$ ,
  - (b)  $\sup_{\{A \in \mathcal{A} : \lambda(A) > 0\}} P(A)/\lambda(A) < \infty$  and
  - (c)  $P(A) = 0$  if and only if  $\lambda(A) = 0$ .

*Proof.* The implication (i)  $\Rightarrow$  (ii) is obvious. If  $A$  and  $d$  are as in (ii) there exists a continuous linear functional  $\phi^A$  on  $L^1(\lambda)$  and  $a$  and  $b$  such that

$$\sup_{x \in \bar{\mathcal{C}}} \phi^A(x) < a < b < \phi^A(d\mathbf{1}_A)$$

Given that  $\mathcal{C}$  contains the convex cone  $-\mathcal{S}(\mathcal{A})_+$ , that  $\mathcal{S}(\mathcal{A})_+$  is dense in  $L^1(\lambda)_+$  and that  $\phi^A$  is continuous, we conclude that  $\sup_{f \in L^1(\lambda)_+} \phi^A(-f) < \infty$  i.e. that  $\phi^A \geq 0$ . It follows from [3, Theorem 2] that  $\phi^A$  admits the representation  $\phi^A(f) = \mu^A(f)$  for some  $\mu^A \in ba(\lambda)_+$ . Moreover,

$$\sup_{\{B \in \mathcal{A} : \lambda(B) > 0\}} \mu^A(B)/\lambda(B) \leq \sup_{\{f \in L^1(\lambda) : \|f\| \leq 1\}} \phi^A(f) = \|\phi^A\| < \infty$$

and  $\sup_{h \in \mathcal{C}} \mu^A(h) < a < b < d\mu^A(A)$  so that  $\mu^A$  meets (a) and (b) above. The inclusion  $0 \in \mathcal{C}$  implies  $a > 0$  so that  $\mu^A(A) > 0$ . By normalization we can assume  $\|\phi^A\| \vee a \leq 1$ . The collection  $\mathcal{M} = \{\mu^A : A \in \mathcal{A}, \lambda(A) > 0\}$  so obtained is dominated by  $\lambda$  and therefore by some  $m \in \mathbf{A}(\mathcal{M})$ , by Theorem 1. Thus  $m \leq \lambda$  and  $\sup_{h \in \mathcal{C}} m(h) \leq 1$ . If  $A \in \mathcal{A}$  and  $\lambda(A) > 0$  then  $m \gg \mu^A$  implies  $m(A) > 0$ . The implication (ii)  $\Rightarrow$  (iii) follows upon letting  $P$  be the finitely additive probability obtained from  $m$  by normalization. Let  $P$  be as in (iii) so that  $L^1(\lambda) \subset L^1(P)$ , by (b). If  $f \in L^1(\lambda)_+$  and  $\lambda(f) > 0$  then  $f \wedge n$  converges to  $f$  in  $L^1(\lambda)$  [6, III.3.6] so that we can assume that  $f$  is bounded. Then, by [1, 4.5.7 and 4.5.8], there exists an increasing sequence  $\langle f_n \rangle_{n \in \mathbb{N}}$  in  $\mathcal{S}(\mathcal{A})$  with  $0 \leq f_n \leq f$  such that  $f_n$  converges to  $f$  in  $L^1(\lambda)$  and therefore in  $L^1(P)$  too. For  $n$  large enough, then,  $\lambda(f_n) > 0$  and,  $f_n$  being positive and simple,  $P(f_n) > 0$ . But then  $P(f) = \lim_n P(f_n) > 0$  so that  $\eta f$  cannot be an element of  $\overline{\mathcal{C}}$  for all  $\eta > 0$  as  $\sup_{h \in \overline{\mathcal{C}}} P(h) < \infty$ .  $\square$

An application of Corollary 3 is obtained in [2, Lemma 3.1]. Corollary 1 also provides a finitely additive version of a useful result of Mukherjee and Summers [8, Lemma 3], illustrating the countable structure of the atoms of an additive set function<sup>1</sup>.

**Corollary 4** (Mukherjee and Summers). *Let  $\lambda$  have atoms. There exists a countable, pairwise disjoint collection  $G_1, G_2, \dots$  of  $\lambda$ -atoms of  $\mathcal{A}$  such that for any  $\lambda$ -atom  $B \in \mathcal{A}$  there exists  $n \in \mathbb{N}$  such that  $\lambda(B \Delta G_n) = 0$ .*

*Proof.* Apply Corollary 1 with  $\mathcal{H}$  the collection of all  $\lambda$ -atoms of  $\mathcal{A}$ . Let  $\langle H_n \rangle_{n \in \mathbb{N}}$  be the corresponding sequence in  $\mathcal{H}$  and put  $G_n = H_n \setminus \bigcup_{i < n} H_i$ . Upon passing to a subsequence if necessary we may assume  $\lambda(G_n) > 0$  so that  $G_n \in \mathcal{H}$  for each  $n \in \mathbb{N}$ . If  $B \in \mathcal{H}$  it follows from (5) that  $\lambda(B \cap G_n) > 0$  for some  $n$ . Given that  $B$  and  $G_n$  are atoms then  $\lambda(B \setminus G_n) = \lambda(G_n \setminus B) = 0$ .  $\square$

## REFERENCES

- [1] K. P. S. Bhaskara Rao, M. Bhaskara Rao: *Theory of Charges*, Academic Press, London, 1983.
- [2] G. Cassese: *Convergence in Measure under Finite Additivity*, Shankyā A, **75** (2013) 171-193.
- [3] G. Cassese: *Sure wins, Separating Probabilities and the Representation of Linear Functionals*, J. Math. Anal. Appl. **354** (2009), 558-563.
- [4] B. de Finetti (1937), *La Prévision: Ses Lois Logiques, ces Sources Subjectives*, Ann. I.H.P. **7**, 1-68.
- [5] J. Diestel, J. J. Uhl Jr.: *Vector Measures*, Mathematical Surveys, N. 15, Am. Math. Soc., Providence, 1977.
- [6] N. Dunford, J. T. Schwartz: *Linear Operators. General Theory*, Wiley, New York, 1988.
- [7] P. R. Halmos, L. J. Savage: *Application of the Radon-Nikodym Theorem to the Theory of Sufficient Statistics*, Ann. Math. Stat. **20** (1949), 225-241.

---

<sup>1</sup>I am in debt with an anonymous referee for calling my attention on this paper.

- [8] T. K. Mukherjee, W. H. Summers: *Functionals Arising from Convergence in Measure*, Amer. Math. Month. **81** (1974), 63-66.
- [9] M. Sion: *On General Minimax Theorems*, Pacific J. Math. **8** (1958), 171-175.
- [10] J. A. Yan: *Caractérisation d'une Classe d'Ensembles Convexes de  $L^1$  ou  $H^1$* , Séminaire de Probabilité XIV, Lecture Notes in Math **784** (1980), 220-222.

UNIVERSITÀ MILANO BICOCCA AND UNIVERSITY OF LUGANO

*E-mail address:* `gianluca.cassese@unimib.it`

*Current address:* Department of Economics, Statistics and Management, Building U7, Room 2097, via Bicocca degli Arcimboldi 8, 20126 Milano - Italy